DOCUMENT RESUME

ED 358 062 SP 034 516

AUTHOR McEneaney, John E.; Sheridan, E. Marcia
TITLE Evaluating the Effectiveness of an Undergraduate Teacher Education Program.
PUB DATE May 93
NOTE 25p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Comparative Analysis; Curriculum Evaluation; Education Courses; Elementary Secondary Education; *Factor Analysis; General Education; *Graduate Surveys; Higher Education; *Measurement Techniques; National Surveys; *Preservice Teacher Education; Program Effectiveness; *Program Evaluation; Special Education; Teacher Attitudes; Teacher Education Curriculum; *Teacher Education Programs; Undergraduate Study

ABSTRACT

This paper describes the development, implementation, and results of a 3-stage model of program assessment for teacher education. A survey instrument of 46 items was developed utilizing Loadman and Gustafson's National Database for Teacher Education. Data were derived from a telephone interview survey of 201 out of approximately 400 recent teacher education graduates of Indiana University at South Bend to assess their perceptions of the adequacy of their training both in general education and in professional preparation programs. The study involved: comparisons of results to those reported in a national survey; analyses to reveal changes across time and demographic dimensions; and a factor analysis of the survey items in an effort to assess the program and refine the conceptual framework of the instrument. Consistent with previous studies, results indicate more positive evaluations from more recent graduates and from graduates who have never taught. In addition, although initial intuition in this study suggested three item categories, the data collected suggested a more informative conceptual framework composed of at least six factors that seem to correspond to curricular blocks of courses (professional skills, general professional knowledge, specialized professional knowledge, humanities, educational foundations, and mathematics/science). The factor structure appears to be well suited to identifying relative strengths and weaknesses in teacher education programs. Findings are displayed in two tables and four bar graphs. (Contains 10 references.) (LL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED358062

COVER PAGE

Evaluating the Effectiveness of an Undergraduate Teacher Education Program

John E. McEneaney E. Marcia Sheridan

Indiana University South Bend

John E. McEneaney is an Assistant Professor of Reading Education with interests in teacher education and literacy acquisition.

E. Marcia Sheridan is Dean of the Division of Education with interests in teacher education and international issues in literacy education.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. McEneaney

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running head: Evaluating Teacher Education

SB034516

BEST COPY AVAILABLE

Evaluating the Effectiveness of an Undergraduate Teacher Education Program

Abstract

This article describes the development, implementation, and results of a three-stage model of program assessment for Teacher Education. Primary data in the research consists of a telephone interview survey of teacher education graduates to assess their perceptions of the adequacy of their training both in general education and in their professional preparation programs. Data analysis involved comparisons of our results to those reported in a national survey, analyses to reveal changes across time and demographic dimensions, and a factor analysis of our survey items in an attempt to assess our program and refine the conceptual framework of our instrument. Our results are similar to previous studies indicating more positive evaluations from more recent graduates and from graduates who have never taught. In addition, although our intuitive assessment of the items used suggested 3 item categories, the data we collected suggest a more informative conceptual framework composed of at least 6 factors that seem to correspond to curricular blocks of courses. The factor structure that resulted thus appears to be well suited to identifying relative strengths and weaknesses in teacher education programs.

Evaluating the Effectiveness of an Undergraduate Teacher Education Program

Curriculum and program review are an important part of the improvement of university instructional programs. Schools and colleges of education must show evidence of follow-up studies as criteria for NCATE professional accreditation by engaging in systematic program evaluations such as follow-up studies. The purpose of this article is to describe our efforts to develop and implement a 3-stage model for program assessment that provides a framework for on-going program evaluation and improvement. In addition we will report on results from our first cycle of data collection, discuss how these results were used in evaluating our instructional programs, and propose a theoretical model of the dimensions of teacher education revealed by factor analytic techniques.

Stage 1: An Arts & Sciences Perspective on Teacher Education

Undergraduates completing teacher preparation programs at IUSB take 48-68% of their credit hours in general education courses with the remainder in professional education. Considering the large number of Arts and Sciences courses that pre-service teachers take, these faculty members play a large part in the education and preparation of teachers for children of today. As a result, we believed that these faculty needed to be involved in this on-going assessment process. In order to understand how general education faculty perceived their role in the education of future teachers, in winter of 1991, a random sample of general education faculty were invited to participate in one of four FOCUS groups and asked the following questions:

1. What should a college graduate know?
2. What should the graduate of an undergraduate teacher education program know?
3. Where should teacher education students attain what they need to know?
4. How can we know if they have attained what is needed?
5. It is now 1995, there is no Division of Education, would you do anything differently if you had the total responsibility for educating teachers in

your areas of expertise?

Group responses were videotaped and analyzed with the following results. General education faculty believed that college graduates needed to know how to think critically and be proficient in basic skills, particularly reading and writing. They believed that graduates from a teacher preparation program should know how to teach, answer questions, convey knowledge and communicate. As to where teacher education students should attain what they needed to know, they believed that methods and teaching skills should be learned from the Division of Education through opportunities for early experiences in the classroom and through modeling. In response to the question how can we know if they have attained what is needed, they stated that students should be receiving continuous feedback and having their performance in the field evaluated. As to the hypothetical question on what they would do differently if there were no Division of Education and they were responsible for educating teachers, they responded that they would recreate the Division of Education. These data were shared with Education and general education faculty, as well as teachers and area school administrators.

Stage 2: Assessing Perceptions of IUSB Teacher Education Graduates

The next step in the process was to look more specifically at how recent IUSB Teacher Education graduates perceived the quality of their undergraduate preparation program. In particular, we wished to look at our graduates' perceptions of their general education and professional education preparation as distinct elements in our assessment process. While standards are set both by the state and NCATE and programs designed to meet these standards, it was the <u>perceived effectiveness</u> of these programs that we sought to determine. As opposed to the more generic questions asked of general education faculty, assessment questions which would lend themselves to specific program improvement were

desired.

The intent to focus on more specific aspects of our program required that we employ a far more specific instrument than that used in Stage 1 of our assessment project. With this problem in mind we turned to the teacher education program assessment literature, hoping to find appropriate models for the development of the instrument we would ultimately use in our study. Initially, the literature review suggested a less than optimistic outlook. Galluzzo and Craig (1990) have, for example, characterized program evaluation as the "orphan" of teacher education. According to their review of the literature, efforts to establish models of program evaluation for teacher education have typically been "site-specific, idiosyncratic models designed for particular applications at individual institutions (p. 602)." Although it can be reasonably argued that the proposal to develop an assessment instrument for evaluating the teacher education program at IUSB was, in fact, site-specific, there were two reasons we wished to adopt a broader view.

One reason for developing an assessment instrument that went beyond immediate "in-house" needs is that, while such an instrument might provide some useful information about graduates' perceptions of the IUSB program, it would not provide for meaningful comparisons to similar studies conducted at other institutions. A second reason for adopting a broader view was that one of the most common criticisms of program evaluation studies cited in the teacher education literature is that they tend to be one-shot studies that are theoretically ill-grounded and generally fail to play a role in a larger on-going evaluation scheme.

We found in our review, however, that there were at least two different approaches to evaluation that appeared relevant to the division's needs. One approach involved adapting an instrument originally developed for use in an

objectives-based teacher assessment program and the other involved the use of a "standardized" assessment instrument especially developed to allow comparisons across institutions.

The concept behind objectives-based evaluation is that teaching competency is reflected in the extent to which teachers demonstrate specific objectives in their teaching. Presumably, the objectives upon which the instrument is based are related to teaching effectiveness. On that assumption, it seems reasonable to think of these objectives as one standard by which teacher preparation programs can be evaluated. In one study (Murray, 1991) similar to the one described here, a random sample of 1985-1990 graduates of the Berry College teacher education program were mailed surveys asking them to report how adequately they felt their education had prepared them for completing the competency-based assessment program for beginning teachers that had been in place in Georgia since 1980. In effect, the Georgia Teacher Performance Assessment Instrument (TPAI) provided the conceptual framework within which graduates were asked to evaluate their professional training. One important limitation of the Berry College study was that only recent graduates were surveyed, yet results reported by Adams (1987) suggest that beginning teachers' perceptions of the adequacy of their undergraduate training appears to be quite unstable until about the fifth year of teaching. Given Adams' findings, we decide to include a years-of-teaching-experience item in the demographic data we collected.

In addition to the TPAI, a variety of other competency-based assessment instruments have been developed for use across the nation. Some of these instruments (like the TPAI) are characterized as high-inference measures that provide numerical scores intended to represent level of competence. Other measures like the Classroom Observations Keyed for Effectiveness Research (COKER), developed at Toledo University are characterized as low-inference,

intended to simply discriminate between competent and non-competent performance in each of the areas assessed. Factor analytic studies (Wiersma et al., 1983) suggest that although the TPAI and the COKER appear to account for the same degree of variance, these two instruments appear to have very different factor structures with the TPAI resulting in two factors whereas the COKER produced five consistently strong factors. Since the goal of our program assessment is to identify specific program strengths and weaknesses, the factor structure of the COKER would appear to more adequately meet our needs but in this case, since practicing teachers are the intended source of data, the COKER is subject to the criticism that it will not pick up on important differences since, presumably, non-competent graduates will not have obtained a teaching position or will have left the profession. The TPAI, by providing a high-inference measure, presumably would allow us to assess degrees of program effectiveness, but its factor structure does not appear to support distinct curricular areas.

A second approach to program evaluation employed a general purpose questionnaire specifically developed to allow comparisons between teacher education programs at different universities. In the study that employed this approach (Loadman, 1989), program evaluators from ten institutions came together at the National Center for Research on Teacher Education at Michigan State University in order to develop a common questionnaire that could be used by a wide variety of institutions so that meaningful comparisons could be made. Preliminary results reported by Loadman and Gustafson (1990) suggested that the questionnaire might help us both conceptualize our assessment program and allow us to make comparisons to programs at other universities.

Our review of the literature suggested that, while we might serve our immediate short-term goals by designing an instrument tailored especially for our curricular areas and students, we would as a result fail to achieve other

important long-term goals related to the ways we conceptualize and implement our program in teacher education and how our program to prepare teachers compares to other institutions across the country. We felt it was in the best interest of our long-term development as a program to participate in the National Database program and as result adapted the short form of the National Database survey for our use.

Stage 3: Adaptation of the survey and research methodology

Adaptation of the National Database survey for our needs involved three modifications. Since our data was collected by phone, items were rephrased, if necessary, to be consistent with that format. Secondly, we adapted the response scale employed in the survey. The original short-form National Database survey employed two different response scales for attitudinal items, a three-point scale for professional skills and a seven-point scale for all other scaled items. We decided to employ a single seven-point Likert scale for all scaled items. Our third modification was to add a number of items to our survey from the long form of the National Database survey. Items 64, 65, 72, 76 and 81 from the long-form survey were added (becoming items 23-27 in our survey) since they addressed professional preparation issues of special interest to us. Items 46-58 from the long form were also added (becoming items 31-43 on our survey) in order to provide a measure of broader subject area knowledge both in professional and general education.

The adapted survey consists of a total of 46 items. The first 10 and last 3 of these items request demographic information, information about the subjects' current professional status and some general attitudinal measures reflecting subjects' general philosophy of learning and teaching. Seventeen items (Items 11-27) were designated SKILL ITEMS since they focus on teaching skills such as selecting, preparing and using media in instruction (Item 23), planning

stimulating lessons (Item 11), and other specifically professional skills. Three items (Items 28-30) were designated QUALITY ITEMS since they asked respondents to rate the overall quality of their student teaching/internship experiences (28), liberal arts/general education courses (29) and courses in their professional program of study (30). Thirteen items (Items 31-43) were designated SUBJECT ITEMS since they focus on graduates' perceived level of knowledge in broader subject areas that include both general and professional education.

Subjects

A list of approximately 400 Teacher Education 1988-1992 program graduates was generated from our files. An attempt was made to contact each individual on the list and administer the survey as a telephone interview with subjects' responses coded directly into electronic files. If after three attempts individuals were not successfully contacted, they were dropped from the subject pool. A total of 201 interviews were completed including 131 Elementary Education, 54 Secondary Education and 16 Special Education program graduates.

Data analysis

Since we had opted to make use of the National Database survey form, our first analyses consisted of comparing our data to that reported by Loadman and Gustafson (1990) and Freeman (1993). Moreover, since means and standard deviations were reported, it was possible to carry out a number of population t-tests (treating the database as the population) in order to assess the likelihood that differences could be attributed to random variation.

In addition to comparisons to the National Database, we were interested in looking at student responses across the five years of graduates included in the survey. If, for instance, there was a declining trend in our graduates' attitudes toward their programs we reasoned that that would be a cause for concern, whereas an improving trend would be viewed more positively. Analysis

of data across time was carried out on three levels.

The first level of analysis across time involved calculating a mean rating across all 33 Skill, Quality and Subject items, resulting in a single overall score that we interpreted as reflecting a global evaluation of our students' educational experience at IUSB. These global means were then analyzed in a one-way ANOVA (Alpha = .05) to assess whether any changes had occurred across the years of interest (1988-1992).

In the second level of data analysis across time we calculated mean scores for each subject within the Skill, Quality and Subject sets of items. We interpreted these scores as reflecting subscale measures for our instrument. Each of these subscale means were then analyzed in a one-way ANOVA (Alpha = .05) to assess whether any changes had occurred across the years of interest (1988-1992).

In the third level of data analysis across time each of the 33 Skill, Quality and Subject items were individually analyzed. Analysis of individual items involved a series of one-way ANOVAs to determine whether there were statistically significant differences in individual item means across years. The purpose of this analysis was to determine what, if any, changes had occurred across time in the evaluation of specific areas in our program by our graduates. Presumably, if specific aspects of our program had improved (or declined) in quality, our graduates' evaluations would be influenced. In addition, if major changes in our program could be related to specific changes in the way graduates responded to our survey we conjectured that we might be in a better position to evaluate the impact of those changes.

Since, in the analysis of individual items we planned multiple comparisons (one ANOVA for each of the 33 items of interest), it became necessary to consider both familywise and per comparison alpha values where the familywise alpha set

an upper limit for type I error across all comparisons and the per comparison alpha set an upper limit for type I error on each comparison. Setting the familywise alpha at a maximum of .05 resulted in a per comparison alpha of approximately .001 where the relationship between these error rates is given by $Alpha_{FW}=1-(1-Alpha_{PC})^{c}$ where c represents the number of planned comparisons (in this case, 33) (Keppel, 1982).

We were also interested to know if teaching experience influenced the way our graduates evaluated their educational experience. Since this data had been collected as a part of our demographics we simply sorted students into groups according to whether or not they had ever taught and reanalyzed our data using average values across all 5 years of graduates in the study.

Finally, we attempted to validate the conceptual structure presumed by our division of items into three categories. Our effort at conceptual validation relied on factor analytic techniques which we used to sort items into categories based on the way our graduates had responded to them on the survey.

Results

Comparison to the national database

Results of the analysis comparing our data to the National Database were quite positive with our graduates reporting significantly higher ratings across Skill ($p<.0000$), Quality ($p<.0000$), and Subject ($p<.0000$) subscales. In addition, comparison to the data reported by Freeman (1993) revealed significantly more positive ratings across nearly all skill items ($p<.003$ in each case), with only items 11 (Planning stimulating lessons) and 25 (Monitoring students' progress ...) resulting in no apparent differences.

Comparisons across time

Results of analyses across time are depicted in Figure 1. Bars represent mean values for all 33 items (AVE) and for each of the three subscales (Quality,

Skill, and Subject). Overall, a steady trend indicating increasingly positive ratings across the years of our graduates is evident.

A one-way ANOVA indicated significant differences across time for our global measure (AVE) averaging all 33 items ($p=.0037$). A post-hoc Scheffe analysis revealed a significant group difference between the 1989 and the 1992 graduates ($p=.0196$). Results of the analysis of the Skill, Quality and Subject subscale scores revealed a significant difference ($p=.0003$) across years for the Skill subscale only. Sheffe post-hoc analyses of the Skill subscale data revealed significant differences ($p<.05$) between 1989 graduates and graduates from 1991 and 1992. Analysis of the Quality and Subject subscores did not indicate significant differences across years.

Results of the individual item data analysis revealed three Skill items (items 13, 26 & 27) with significant increases of mean scores across time. In addition, the means of two other Skill items (11 & 25) showed marginal increases across time ($Alpha_{PC} = .003$, $Alpha_{PW} = .09$).

Comparisons by whether or not graduates had taught

Results of the analysis by whether or not respondents had ever taught are depicted in Figure 2. This analysis also revealed differences. Significantly more positive ratings by non-teachers for the overall average ($p=.0016$) and the Skill subscale average ($p=.0003$) were evident but no difference was apparent for the Quality subscale ($p=.4074$) and only a marginal difference was apparent for the Subject subscale ($p=.0405$).

Factor analysis

Following preliminary analysis of the data within the conceptual framework that had been presumed, we carried out a factor analysis of all 33 items. A principal components analysis resulted in 9 factors with eigenvalues greater than one accounting for 66% of the total variance. These nine factors were then

subjected to an oblique rotation since we did not expect that the factors of interest would be independent. Of the 9 factors identified (See Table 1) only the first 6 were clearly interpretable. Based on item loadings we labelled these 6 interpretable factors 1) Professional skills, 2) General Professional Knowledge, 3) Specialized Professional Knowledge, 4) Humanities, 5) Educational foundations, and 6) Mathematics/Science.

Numerous items from the Skills subscale loaded heavily on Factor 1 (Professional skills). Nearly all of the items loading heavily on Factor 2 (General professional knowledge) tapped general knowledge having to do with historical, legal, and multicultural issues in education, content typically addressed in survey courses in our program. Items having to do with individualized learning, special education and communicating with parents loaded heavily on Factor 3 which was labelled Specialized Professional Knowledge. Since the survey was intended to assess general education outcomes (as well as professional education) we were not surprised to find that content typically delivered in courses our students take in the Division of Liberal Arts and Sciences fell out in a separate factor (Factor 4) which we labelled the Humanities factor. Items loading most heavily on Factor 5 were those that dealt with educational theory (child development, learning theory), content typically addressed in our educational foundations block. Finally, items 31 and 32 dealing with mathematics and science fell out in Factor 6.

Since the rotation employed was oblique, the analysis also resulted in a set of inter-factor correlations (See Table 1.) that appears to support our interpretation of the factors. Factor 1 correlates fairly highly with Factors 2 and 3 but none of the other factors correlate highly with one another.

Following this factor analysis we reanalyzed our data using the conceptual structure that had emerged. New subscale scores for each of the 6 interpretable

factors were calculated for each student. New analyses by year of graduation and whether or not the respondent had taught were carried out. Results of these reanalyses depicted in Figures 3 and 4 and Table 2 reinforced our earlier findings but also suggested more subtle effects we had not seen before. Factor 1 (Professional skills) showed significant differences across time in the expected direction. (More recent graduates reporting more positive evaluations.) The Specialized Professional Knowledge factor showed a marginal difference ($p=.0417$) across time. Analysis according to whether subjects had taught revealed significant differences both for the Professional Skills factor ($p<.0000$), the Humanities factor ($p=.0021$), and the General Professional Knowledge factor ($p=.0288$). The Specialized Professional Knowledge factor again showed marginal changes across groups.

Discussion

Two results emerge from our analysis that are of greatest interest to us. One result is that, consistent with a number of other previous studies (Gaede, 1978; Ligana, 1970; Hummel & Strom, 1987), more recent graduates appear to have a more favorable evaluation of their educational experience than do graduates with more teaching experience. A careful analysis of the data suggests, however, that most of the variation across years can be attributed to changes in the way graduates responded on the Skill subscale. Moreover, item-by-item analysis of the complete set of 16 skill items suggests that much of the variation across time can be attributed to a relatively small number of items within the Skill subscale. Specific items for which significant differences were found include:

1) Teaching problem solving and higher-order thinking skills (Item 13, $p=.001$),
2) Reflecting and improving your teaching performance (Item 27, $p=.000$), and
3) Using jigsaw, TGT (teams-games-tournament), and other cooperative learning techniques (Item 26, $p=.000$).

Two additional items showed increased means across time at an alpha level slightly above that which was set prior to analysis.

3) Monitoring students' progress and adjusting instruction (Item 25, p=.003).

4) Planning stimulating lessons (Item 11, p=.003).

Several possible explanations can be offered for the results obtained. One possible explanation is that the more positive responses of more recent graduates is simply a halo effect that tarnishes with time, an effect Ligana (1970) has termed the "Curve of disenchantment." It may be that graduates leave the program feeling quite positive about their educational experience but that, with exposure to the difficulties of teaching in real classrooms, they begin to feel less positively. Arguing against this interpretation, however, is the fact that only 3 of 33 items on the survey showed significant changes across time and only 2 more showed marginal changes. In addition, all of those changes occurred in the Skills subscale. Presumably, if changes across time were due to a more general disenchantment we would expect to see more consistent changes across our items across all of our subscales rather than only a few in a single subscale.

The suggestion that survey responses actually reflect program characteristics is further supported by changes instituted in our program over the years investigated in the study. One of the program changes that occurred at about the midpoint of our survey sample was an increased emphasis on cooperative learning techniques in the division. A graduate workshop especially devoted to this topic was started and undergraduate students began to receive more instruction in this area in their general and specific methods courses. Inspection of year-by-year means for this item reveals a substantial jump at about the midpoint of the survey span. Another important influence on our program during the years studied was our effort to clarify the roles of decision making and reflective evaluation as components of our teacher education model.

The increased emphasis in these areas was also reflected in the responses of our graduates. It therefore appears that program changes are related to our survey results in ways we would expect, thus strengthening the argument that this kind of data can provide insights on program characteristics and the effects of programmatic changes.

Our second major finding is that, although our initial categorization of Skill, Subject and Quality items helped us organize our assessment effort, the data we collected suggest a more informative conceptual framework composed of at least 6 factors that seem to correspond to curricular blocks of courses. This is especially important for us since it was program assessment that drove this study from the beginning. The factor structure that resulted thus appears to be well suited to our interests in identifying relative strengths and weaknesses in our program.

Summary and Conclusions

This paper describes a three-stage process of program assessment developed and implemented in the Division of Education at IUSB over the course of the past year and a half. Initially, Arts and Sciences faculty participated in FOCUS groups intended to clarify an "outside" perspective on the needs of teacher education students and the best way to provide for those needs. In the second stage of the assessment process we turned to a telephone interview survey of recent graduates to assess their perceptions of the adequacy of their training both in general education and in their professional preparation programs. In the third stage of the process we factor analyzed our survey items in an attempt to refine its conceptual framework.

What is perhaps most encouraging about our survey results at the conclusion of this first cycle of data collection is the finding that documented program changes appear to be reflected in the responses of our graduates. Specific

program modifications that occurred during the years of our survey span appear to have resulted in more positive responses in those areas, thus supporting the intended use of this kind of survey data. It has also been encouraging to find that the factor analysis of our data suggests that students perceive their relative strengths and weaknesses in terms of meaningful curricular areas since this makes interpretation of survey results and the application of these results to program improvement much more straightforward.

Our intention is to continue to look for relationships between survey results and specific program characteristics or changes that we can document from other sources. We also intend to rank order all of the scaled items on the survey in order to identify those on which graduates responses were most positive and most negative. We expect to continue our program assessment based on this data by asking why our graduates have rated these items the way they did. Can we find any explanations for their responses in what we know about our program? In other words, is there any other evidence to support these areas as relative strengths and weaknesses? Finally, as has been suggested, it is our intent to make this kind of survey a regular part of an on-going assessment program which will allow us to better understand what, if any, consideration must be given to the stability of response factor noted by Adams (1987) and Ligana's (1970) "Curve of disenchantment" described above, both of which are related to recency of graduation and thus could have implications for interpreting our results.

References

Adams, R. D. (1987). Follow-up studies of teacher education graduates. In M. Haberman & J. M. Backus (Eds.) Advances in Teacher Education, (Vol. 3, pp. 181-201). Norwood, NJ: Ablex.

Freeman, D. (1993). Are self-perceptions of competence in teaching a unidimensional or multidimensional trait? Paper presented at the annual meeting of the American Association of Colleges for Teacher Education. San Diego, CA. February, 1993.

Gaede, O. (1978). Reality shock: A problem among first year teachers. The Clearinghouse, 51, 405-409.

Galluzzo G. R. & Craig, J. R. (1990). Evaluation of Preservice Teacher Education Programs. In Handbook of Research on Teacher Education. New York: Macmillan.

Hummel, T. & Strom, S. (1987). The relationship between teaching experience and satisfaction with teacher preparation: A summary of three surveys. Journal of Teacher Education, 38(5), 28-36.

Ligana, J. (1970). What happens to the attitudes of beginning teachers? Danville, IL: Interstate.

Loadman W. E. (1989). Developing a national database for preservice teacher education follow-up studies. Paper presented at the meeting of the American Association of Colleges of Teacher Education, Anaheim, CA.

Loadman, W. & Gustafson, G. (1990). National Database for Teacher education Program Follow-up. 1990 Annual Report. Columbus, OH: The National Database for Teacher Education Program Follow-up.

Murray, D. (1991). An analysis of the perceptions of teacher education graduates on their preparation for service: A five year review. (ERIC document Reproduction Service No. ED 335 349.)

Wiersma, W., et al. (1983). Assessment of teacher performance: Constructs of teacher competencies based on factor analysis of observation data. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, April 11-15, 1983. (ERIC document Reproduction Service No. ED 230 586.)

Table 1. Inter-factor correlations for the 6 interpretable factors that emerged following an oblique (Promax) rotation.

	FACTOR1	FACTOR2	FACTOR3	FACTOR4	FACTOR5	FACTOR6
FACTOR1	100 *	43 *	42 *	28	30	24
FACTOR2	43 *	100 *	34	27	16	25
FACTOR3	42 *	34	100 *	25	11	13
FACTOR4	28	27	25	100 *	12	16
FACTOR5	30	16	11	12	100 *	4
FACTOR6	24	25	13	16	4	100 *

NOTE: Printed values are multiplied by 100 and rounded to the nearest integer. Values greater than 0.387986 have been flagged by an '*'.

Table 2. Results of analyses of variance for Factors 1-6 across years of graduation and whether or not the respondent had ever taught.

	Changes across year of grad.	Teach/Nonteach Differences
Factor 1 (Professional Skills)	**Y (p<0.0001)***	**Y (p<0.0000)****
Factor 2 (Gen. Prof. Knowledge)	N (p=0.2522)	**Y (p=0.0288)****
Factor 3 (Spc. Prof. Knowledge)	**Y (p=0.0417)***	**Y (p=0.0411)****
Factor 4 (Humanities)	N (p=0.2506)	**Y (p=0.0021)****
Factor 5 (Ed. Foundations)	N (p=0.1475)	N (p=0.4254)
Factor 6 (Math/Science)	N (p=0.9677)	N (p=0.5949)

* More recent graduates reporting more positive evaluations.

** Non-teaching graduates reporting more positive evaluations.

Figure 1. Mean ratings by undergraduates of their training (1=Poor, 7=Excellent) for all items (Average), and for Quality, Skill, and Subject subscales across five graduating classes (1988-1992).

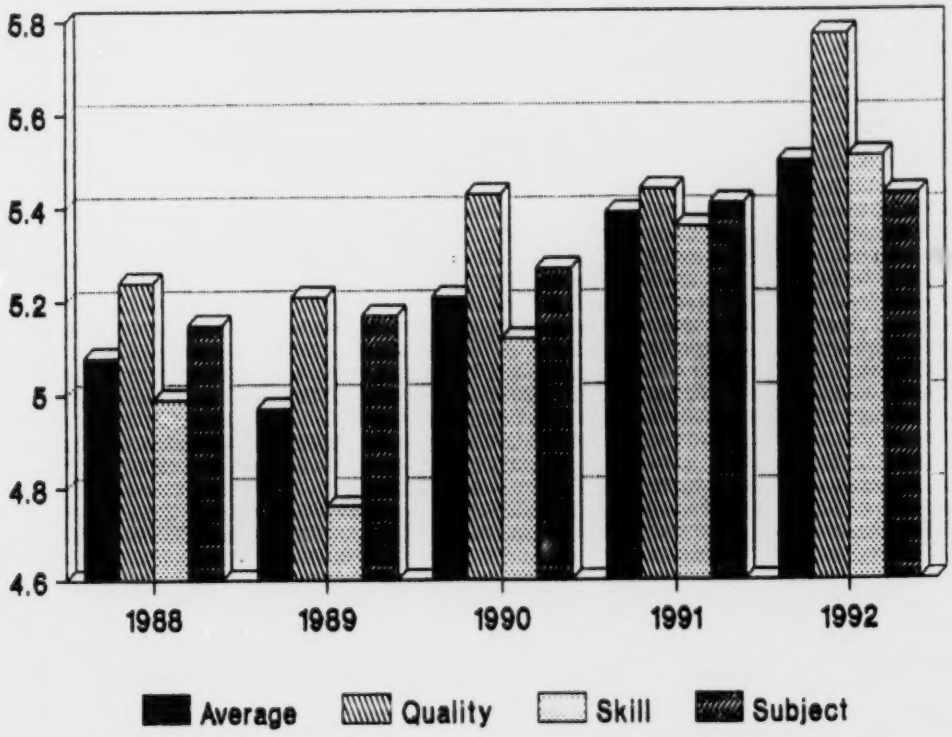

Figure 2. Mean ratings by undergraduates of their training (1=Poor, 7=Excellent) for all items (Average), and for Quality, Skill, and Subject subscales according to whether or not they had ever taught.

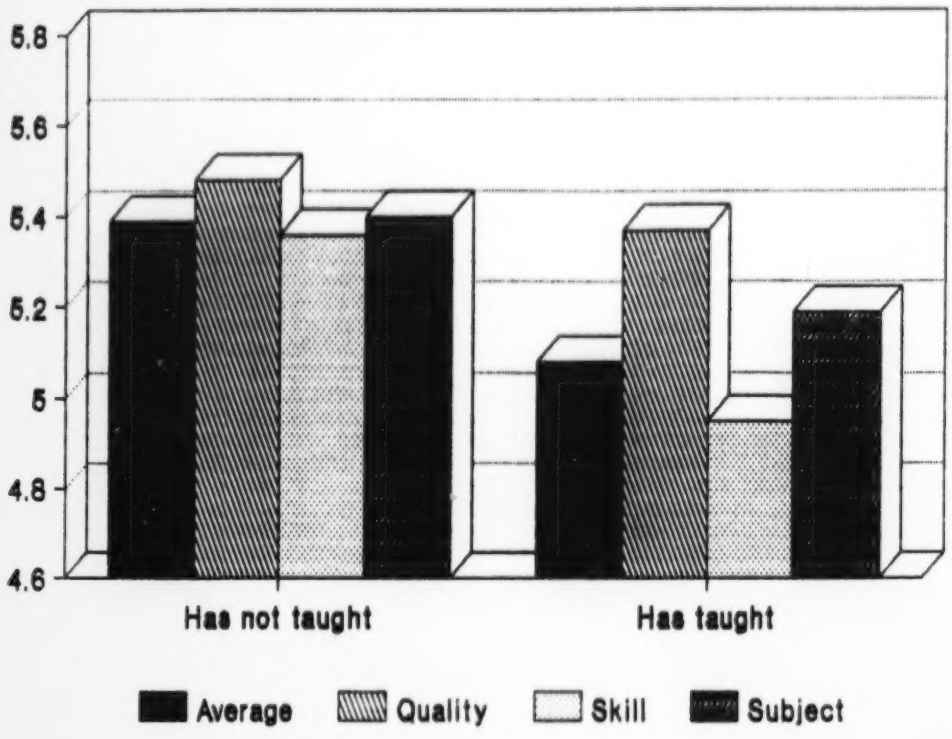

Figure 3. Mean ratings by undergraduates of their training (1=Poor, 7=Excellent) for each of the 6 factors identified across five graduating classes (1988-1992).

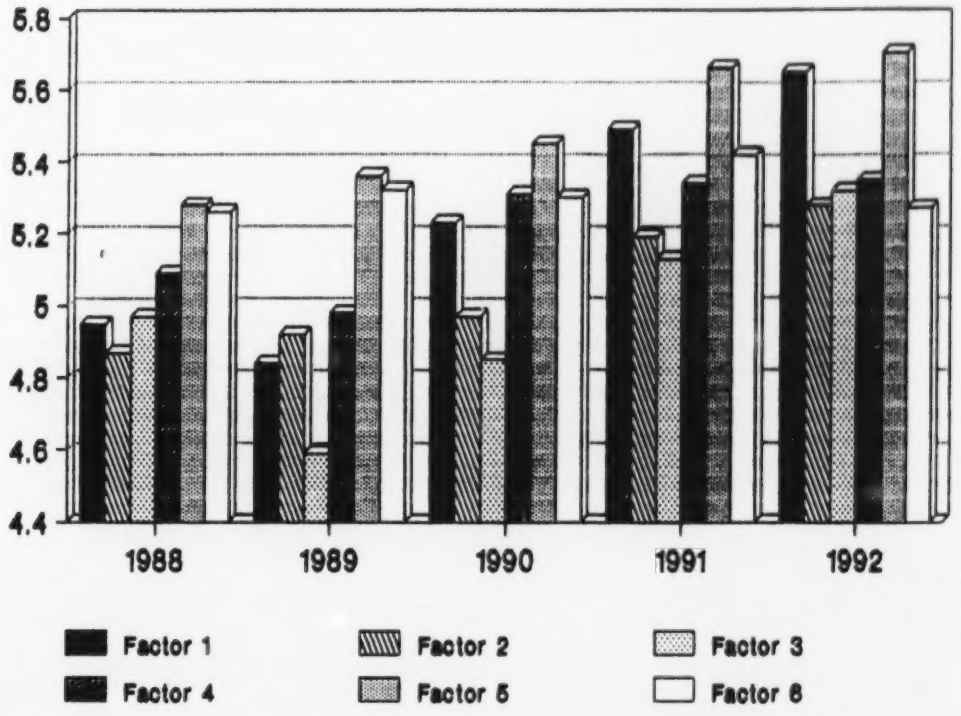

Figure 4. Mean ratings by undergraduates of their training (1=Poor, 7=Excellent) for each of the 6 factors identified according to whether or not they had ever taught.